Low Power CMOS Vision Sensor for Gaussian Pyramid Extraction

M. Suárez, V.M. Brea, J. Fernández-Berni, R. Carmona-Galán, D. Cabello, and A. Rodríguez-Vázquez *Fellow, IEEE*

Abstract

This paper introduces a CMOS vision sensor chip in standard 0.18 μm CMOS technology for Gaussian pyramid extraction. The Gaussian pyramid provides computer vision algorithms with scale invariance, which permits to have the same response regardless of the distance of the scene to the camera. The chip comprises 176 $\times$ 120 photosensors arranged into 88 $\times$ 60 processing elements (PEs). The Gaussian pyramid is generated with a double-Euler switched-capacitor network. Every processing element comprises four photodiodes, one 8-bit single-slope Analog to Digital Converter (ADC), one Correlated Double Sampling (CDS) circuit, and 4 state capacitors with their corresponding switches to implement the double-Euler switched-capacitor network. Every processing element occupies 44 $\times$ 44 μm^2. Measurements from the chip are presented to assess the accuracy of the generated Gaussian pyramid for visual tracking applications. Error levels are below 2% full scale output (FSO), thus making the chip feasible for these applications. Also, energy cost is 26.5 nJ/px at 2.64 Mpx/s, thus outperforming conventional solutions of imager plus microprocessor unit (MPU).

Keywords

CMOS Vision Sensors, Gaussian Filters, Image Pyramids, Switched-Capacitor Circuits, Per-Pixel Processing

---

This work has been funded by ONR N00014-14-1-0355 and Spanish government projects MINECO TEC2015-66878-C3-1-R&TEC2015-66878-C3-3-R, Junta de Andalucía, Proyectos Excelencia- Conv. 2012 TIC 2338, EM2013/038 (FEDER), EM2014/012. The authors kindly acknowledge the Univ. of California for the Software Transfer Agreement UC Case No. 2014-453 and doctor Steffen Gauglitz for his assistance with the software framework to extract metrics for visual tracking.

M. Suárez is with Atomos GmbH, Villingen, Germany. Contact: manuel@atomos.com.

V.M. Brea and D. Cabello are with the Centro Singular de Investigación en Tecnoloxías da Información (CITIUS), University of Santiago de Compostela, SPAIN. Contact: victor.brea@usc.es.

R. Carmona-Galán is with Instituto de Microelectrónica de Sevilla (IMSE-CNM), CSIC, SPAIN.

J. Fernández-Berni and A. Rodríguez-Vázquez are with the University of Seville, SPAIN, and with Instituto de Microelectrónica de Sevilla (IMSE-CNM), CSIC, SPAIN. Contact: angel@imse-cnm.csic.es.

I. INTRODUCTION

The integration of camera systems for vision applications benefits from performing scene analysis right at the sensor front-end. Such pre-processing may extract scene features and hence reduce the number of data transmitted off the sensor chip for further processing. This is quite a relevant characteristic because images contain many spare data, and data transmission and storage consume significant energy and area. Also, pre-processing and reduced data transmission result in increased throughput. Actually, pre-processing is smartly implemented in natural vision systems [1], [2]; a fact that has motivated authors to explore architectures for CMOS imaging front-ends with per-pixel processing circuitry [3]–[7]. These systems are recently making the transition from academic proof-of-concept prototypes to industrial products [8].

Sensory-processing front-end chips with per-pixel processors operate typically as Single Instruction Multiple Data (SIMD) processors, namely, all processors run concurrently the same operation on the data captured by the pixel photosensors, thus accelerating computation. Also, mixed-signal per-pixel processors provide speed advantages with large energy efficiency [9], [10]. As a result, image sensors with embedded mixed-signal processors emerge as suitable candidates for the front-end of vision systems with optimum SWaP (Size, Weight and Power) figures and large throughput. Throughout the paper we will use the term CVIS (CMOS VIsion Sensors) to refer to image front-end devices with embedded analysis capability and, we will retain the term CIS (CMOS Image Sensors) for conventional image front-ends conceived to deliver just images.

Major points hampering further development of CVIS-SIMD are: i) their outcome may not be compatible with computer vision software tools, thus limiting their acceptance by system engineers and integrators; ii) reduced fill-factor when realized in standard 2D technologies; iii) large pitch, and hence smaller resolution than CIS per given form factor, again in standard 2D technologies. Nevertheless, the loss of resolution and image quality of CVIS-SIMD are not insurmountable barriers for vison. Nature also teaches lessons in this regard; for instance, patients with retinitis pigmentosa see with a small fraction of their photoreceptors alive [11], which suggests that large pixel counts may not be a must. Indeed, resolutions as low as 32 $\times$ 32 pixels suffice to get the gist of complex scenes [12] and have been demonstrated for indoor elderly care [13]. Also, commercial sensors with low pixel counts (QCIF: 176 $\times$ 144) are produced for machine vision applications [14] and have been demonstrated for adaptive laser welding [15], among other applications. Also, reduced fill-factor may be overcome with controlled illumination, as it actually happens in many machine vision applications [16]. Furthermore, many computer vision algorithms cope with inaccuracies arisen during processing [17], [18], thus easing the use of mixed-signal CVIS-SIMD. As an example, the chip in [19], that runs the earliest stages for face detection using the algorithm in [17], tolerates processing errors close to 10%. As shown in Section IV.D, chip measurements in this paper show that inaccuracies in the Gaussian pyramid are low enough as not to be a concern for visual tracking.

Regarding compatibility with computer vision tools, it can be met by aligning the conception of CVIS-SIMD to standard computer vision procedures [20]. Particularly, by focusing on the embedding of pre-processing functions customarily used by computer vision system engineers. This is actually the case of image pyramids, such as the Gaussian pyramid [21]. Image pyramids are found at the initial stages of the processing vision chain for a large variety of computer vision applications and algorithms such as the Scale Invariant Feature Transform (SIFT) and variations thereof. Their calculation is resource intensive because it involves repetitive operations with the whole set of image data. As a consequence, the potential benefit of calculating them with CVIS-SIMDs is huge. CVIS-SIMDs may represent a first step towards embedding complete computer vision on a single die with vision capabilities into SWaP sensitive systems such as vision-enabled wireless sensor networks [22] or unmanned aerial vehicles [23].

From now on we will use the acronym PE (Processing Element) for the elementary cell of CVIS-SIMD image front-end chips. This paper reports a 0.18 μm CMOS sensory processing chip to extract the Gaussian pyramid with per-pixel processing circuitry, ADC and Correlated Double Sampling (CDS). It contains 176 $\times$ 120 3T APSs arranged into 88 $\times$ 60 PEs; i.e.

four photosensing points per-PE. Gaussian filtering is realized by using a diffusive, double-Euler, switched-capacitor grid. The chip operates at 2.64 Mpx/s with an energy consumption of 26.5 nJ/px (0.6 μJ/frame), thus outperforming conventional architectures of imager and MPU by several orders of magnitude. Measurements show errors below 2% FSO versus Gaussian pyramid computed by software [24]; these errors are tolerated by vision applications.

II. GAUSSIAN PYRAMID EXTRACTION

A. Basic Concepts

The scale-space enables computer vision algorithms to give the same response regardless of the distance between camera and object. A common function for scale-space generation is the Gaussian filter [25], [26]. The scale-space is a function $L(x, y, \sigma)$ resultant from the convolution of a variable-width Gaussian function with an input image $I(x, y)$:

$$L(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} * I(x, y) \tag{1}$$

where $*$ is the convolution operator, σ is the width of the Gaussian function, and x, y are the spatial coordinates of the image.

The Gaussian pyramid, illustrated in Fig. 1, consists of several scale spaces arranged into octaves. Starting from the bottom, images within each new octave have all one quarter the resolution of those in the previous octave. Subsampling is hence made in the transition from each octave to the next one. Regarding images contained within each octave, these images are scales obtained through Gaussian filtering with increasing widths. The width of each new scale is k times larger than that of the previous one. The range of scale widths is the same for all octaves, namely, from σ_0 to $2\sigma_0$. The width σ_0 is application-dependent, and as such it could be selected by the user. Usually three octaves with six scales each suffice [21]. At hardware level, the issue is to provide accurate widths σ_i of the Gaussian function.

B. Hardware Implementation

The Gaussian function gives the value I_{ij} of each pixel as the solution of a first-order differential equation under the driving force of the values of the four neighboring pixels along the cardinal directions, namely,

$$\frac{dI_{i,j}}{dt} = D(I_{i+1,j} + I_{i-1,j} + I_{i,j+1} + I_{i,j-1} - 4I_{i,j}) \tag{2}$$

which is actually the continuous-time heat differential equation [27], with D being the diffusion coefficient, usually a constant value common to all the pixels in the image space. In the case of the Gaussian pyramid, D determines the degree of blurring through the expression $\sigma = \sqrt{2Dt}$, where variable t is the time. In our case, pixel values are voltages V_{ij} held at state capacitors of capacitance C, and pixels are connected to the four neighbors through resistive links with resistance R. In such a case, Eq. (2) transforms into,

$$C\frac{dV_{i,j}}{dt} = \frac{V_{i+1,j} + V_{i-1,j} + V_{i,j+1} + V_{i,j-1} - 4V_{i,j}}{R} \tag{3}$$

from where $D = 1/RC$, and the filter width $\sigma_{RC} = \sqrt{2t/RC}$. Resistance R can be implemented either through TMOS transistors operating in ohmic region, giving rise to RC networks, or through switched-capacitor (SC) networks. Fig. 2 illustrates both implementation styles. The former are inherently more non-linear than the latter. Also, RC networks need sampling mechanisms to stop the transient evolution of the network and thereby set the width [4]. The non-linearity of active resistive links and the time uncertainty of sampling mechanisms degrade the accuracy of the diffusion process in RC networks. These problems can be overcome by emulating resistive links through switched-capacitors, giving rise to the so-called diffusive SC networks.

There are many different SC topologies to run Gaussian filters [28]. Fig. 2(b) and 2(c) display simple- and double-Euler SC networks in 1D. In both cases an exchange capacitor C_E is sampled by two switches driven by two non-overlapping clock signals ϕ_1 and ϕ_2 (Fig. 2(d)). The Gaussian pyramid provided by the double-Euler configuration yields better figures of merit than those of the simple-Euler SC topology when included in the SIFT algorithm [29]. Hence, the double-Euler is the SC network implemented on the CVIS-SIMD presented in this paper.

Assuming, as in any SC circuit, that transients associated with the ON resistances of the switches are neglected, that all state capacitors have the same capacitance C, and that C_{E1} = C_{E2} = C_E, the equivalent impedance of the double-Euler SC topology is $R = T_{clk}/nC_E$, where n is the number of clock cycles, and T_{clk} is the clock period. The resultant σ_{SC}, the Gaussian width of the double-Euler SC topology across the number of clock cycles, becomes:

$$\sigma_{SC} = \sqrt{\frac{2nC_E}{C}} \qquad (4)$$

Eq. (4) can be used to set the Gaussian width by design. However, deviations may be observed during fabrication that depend on the actual device employed to implement C_E and C. It is hence convenient to extract the on-chip σ_{SC} value through measurements. Extracted values might be used for calibration if needed. The extraction procedure of on-chip σ_{SC} for our chip will be addressed in Section IV-B.

III. Chip Design

A. Chip Floorplan and Processing Elements

The micrograph at the left in Fig. 3 shows the chip floorplan, consisting of a core array of PEs surrounded by a split frame buffer. The core array includes 88 $\times$ 60 PEs. Each PE comprises: $i)$ 4 3T-APS pixels - spatial resolution regarding image acquisition is hence 176 $\times$ 120; $ii)$ a comparator for in-PE A/D conversion; $iii)$ 4 state capacitors, and a CDS circuit, which is also used as part of Local Analog Memories (LAMs) to store either the acquired scene or a given scale across the Gaussian pyramid, and; $iv)$ the double-Euler SC network made up of intra and inter-PE switches for NEWS connectivity. The inset at the right of Fig. 3 is a close-up of the PEs, where photodiodes and capacitors of the double-Euler diffusion network are visible.

Per-PE ADC and per-PE CDS, instead of the conventional per-column approach, increase parallelism. Also, this strategy gets favoured by the re-targetting of the herein proposed architecture to vertical technologies, leading to better performance metrics [30], [31]. Circuit sharing through the use of the same devices for different functions along time in part compensates for the per-PE ADC and CDS area overhead. Larger routing from the per-PE and per-CDS is alleviated by laying down the frame buffer that stores the results from the A/D conversion in two halves at the top and bottom of the PE array, which in turn diminishes power consumption.

B. PE Array Configuration

The PE array changes its configuration according to the function realized by the chip. Fig. 4 conveys such configurations. The coordinates in the PE array are indicated within brackets. The origin of the coordinates is the PE at the top left corner. State capacitors of the double-Euler SC network in every octave (O_k) are expressed as C_{pij_Ok}.

The input image and the scales in the first octave are stored at state capacitors (C_{pij_O1}). As seen in Fig. 4(a) and Fig. 4(b), as there is only one ADC and CDS circuit per 4 pixels and 4 state capacitors, image acquisition and scales read-out are performed for 4 cycles. State capacitors are shunted across octaves. Fig. 4(c) shows the configuration during the second octave. In this case, the state capacitors of a PE are combined into only one to perform downscaling, which leads to one-to-one state capacitor per CDS and A/D circuit in the PE array. In the third octave, the state capacitors of 4 PEs are merged, and again there is a one-to-one state capacitor per CDS and A/D circuit. The read-out of the input image and the 18 scales resultant from 3 octaves and 6 scales each amounts to 40 A/D conversions of the PE array for the whole Gaussian pyramid.

C. Circuit Implementation

Fig. 5 shows a circuit view of the PE with its time diagram. Table I lists the sizes of the transistors in Fig. 5. Switches are implemented with NMOS transistors with minimum dimensions. Circuit sharing is performed with amplifier $A1$, capacitors C and C_{pij}. Every 3T-APS pixel has its corresponding capacitor C_{pij}. This is shown in Fig. 5 with the same gray color. Capacitor C runs CDS and offset-compensation comparison during A/D conversion. Amplifier $A1$ and capacitors C_{pij} are part of LAMs and CDS circuits. The latter are also part of the state capacitors C_{pij_Ok} in the SC network.

The gain stages in the PE are double-cascode topologies. Only one amplifier is included for CDS and image storing in the LAMs, while two are required in the comparator of the A/D converter. The amplifier can be configured in two modes of operation, namely IA and IB, shown in Fig. 6(a) and Fig. 6(b), respectively. In both cases the current can be cut off through $enable$ ports. Switches driven by $enable$ ports increase their output impedance close to the end of the operating range of the amplifier, increasing the gain too. Configuration IB consumes up to 30% less power than IA at the cost of a narrower input range by shunting the port $enable\_n$ to the input voltage V_{in} (Fig. 6(d)). The bias current of both configurations is set to 1 μA by V_{bp} through a wide-swing constant transconductance bias circuit trimmed with an external resistor [32], leading to a gain above 60 dB in the voltage range [0.4, 1.3] V with mismatch and Process-Voltage-Temperature (PVT) variations (Fig. 6(c) collects nominal simulations). Bode plots are shown in Fig. 6(e).

1) Image Acquisition: The photodiode is an n-well over p-substrate structure in orden to enhance the spectral response at longer wavelengths. The bias current of the source follower of the 3T-APS is set to 1 μA by $M4$ through a transconductance circuit with an external resistor. CDS is included to diminish reset noise and FPN from mismatch [33]. The nominal working range for the output voltage of the CDS circuit is defined by amplifier $A1$ in Fig. 5, namely; [0.4, 1.3] V. These are the lower and upper bounds for the voltages at the state capacitors of the double-Euler SC network.

Fig. 7 shows the CDS topology with its control signals. A similar implementation has been used for instance in [34]. For a given pixel ij, signal ϕ_{rw_pij} is high during the whole acquisition time. Reset and signal voltages for CDS are sampled at time instants t_0 and t_1 with signal ϕ_{acq} high. The

CDS output is stored in C_{pij}, as well as in C'_{pij} and the four exchange capacitors C_E connected to the node n_{ij}. Signals $\phi_{1\_O1\_pij}$ and ϕ_{1_pij} set the initial values in the exchange capacitors used for intra-PE and inter-PE connections in the double-Euler SC network, respectively.

The CDS is implemented with amplifier $A1$ in IA mode to support a wide input voltage range. Enable signal ϕ_{en_inv1} allows switching off amplifier $A1$ between the two samples at t_0 and t_1. By assuming large enough gain $A1$, the CDS output voltage is given by:

$$V_{outij} = V_{ref} + \frac{C}{C_{pij}}[V_{Pij}(t_0) - V_{Pij}(t_1)] \tag{5}$$

where V_{ref} = 400 mV.

2) Local Analog Memories (LAMs): The LAMs store both the image after CDS and the scales across the Gaussian pyramid. The LAMs are implemented with amplifier $A1$, capacitors C_{pij}, and switches ϕ_{writep}, ϕ_{rdm} and ϕ_{write0} (see Fig. 5). Scales across the Gaussian pyramid are stored and read out in two phases with signal ϕ_{rw_pij} high and ϕ_{vref_cds} low. Both phases are shown in Fig. 8. During the first phase voltage $V_{nij} - V_Q$ is held in capacitor C_{pij} with signal ϕ_{rdm} high, and ϕ_{writep} and ϕ_{write0} low. The read-out is performed during the second phase with ϕ_{rdm} low and ϕ_{write0} and ϕ_{writep} high, leaving $V_{outij} = V_{nij}$, where V_{nij} is the voltage at node nij.

3) Comparison for in-PE ADC: Our chip embeds an 8-bit single-slope in-PE ADC. Fig. 9 shows the single-input offset-compensated comparator of the in-PE ADC. Offset-compensation makes the comparator less sensitive to manufacturing variability. Switches are implemented with NMOS transistors. Their sizes are collected in Table II. Label $M15$ means the four transistors in the NAND gate of the comparator, which is implemented with complementary logic. Amplifier $A2$ is configured as IA, while $A3$ is in mode IB to cut power consumption; further decreased with the feedback loop between both gain stages. The bottom sampling technique is run with different delays between signals ($Delay1$ - $Delay3$ in Fig. 9).

The comparator works in two phases: reset and comparison. During reset, both the first input signal and the quiescent point of the first amplifier in the comparator are sampled. This is done with signals ϕ_{comp_rst} and ϕ_{write} high. The reset phase ends by setting ϕ_{comp_rst} and ϕ_{write} low, leaving V_Q-V_{outij} across C. V_{outij} can be either the input image with CDS or a given scale of the Gaussian pyramid. This voltage is compared to the voltage ramp V_{ramp} during the comparison phase, which starts with ϕ_{comp} and ϕ_{ramp_read} high, giving Eq. (6) at the output of the second gain stage. The static power consumption can be cut during reset with ϕ_{comp} low and $\phi_{\overline{en\_comp}}$ high. The comparator takes a falling ramp as input in the comparison phase with a downfall Δ of signal V_{ramp} at V_{OH} = 1.3 V to ensure a correct initial state for values of V_{outij} close to V_{dd}.

$$V_{out_2} = K^2(V_{ramp} - V_{outij}) + V_Q \tag{6}$$

The V_{outij} - V_{ramp} crossing triggers the signal End-of-Conversion (EoC) to low, enabling the writing of a digital word given by an 8-bit counter into the frame buffer assigned. The end of conversion occurs with V_{out2} low (see Fig. 9 and Eq. (6)), which in turn cuts off current in the first gain stage through a positive feedback loop. The feedback loop also reinforces logic levels. Voltage and current waveforms in the first amplifier of the comparator (V_{out1} in Fig. 9) with and without feedback loop plotted in Fig. 10(a) confirm this statement. Fig. 10(b) and (c) illustrate power savings from the feedback loop for two input voltages, corresponding to ADC output codes 250 and 40, close to the lower and upper parts of the falling ramp. Blue and pink lines are the currents integrated along the whole ramp in the first and second amplifiers of the comparator. The comparator without feedback loop consumes 1.65 μW and 1.7 μW for codes 250 and 40, respectively; the feedback loop leads to 75 nW and 1.65 μW, resulting in large power savings for the largest ADC output codes.

4) Gaussian Pyramid Construction: Our double-Euler SC network with NEWS connectivity yields the Gaussian pyramid. Intra- and inter-PE connections are shown in different gray colors in Fig. 5. Fig. 11 gives a complete view of both intra- and inter-PE connections.

Downscaling across octaves in the Gaussian pyramid leads to three types of switching blocks in the SC network, labeled SC_A, SC_B and SC_C in Fig. 11, all of them implemented as NMOS transistors with minimum dimensions. In addition, one out of four PEs has a slightly different structure from the other three. Such a PE is shaded and marked with β in Fig. 11. PEs of α type comprise switching blocks SC_A and SC_B. PEs of β type contain switching blocks SC_A and SC_C. The scales are provided by capacitors C_{pij_Ok}. C_{pij_O1} means any of the 176×120 state capacitors in the first octave. Similarly, C_{pij_O2} and C_{pij_O3} mean any state capacitor in the second and third octaves, where the resolution is downscaled to 88×60 and 44×30 pixels, respectively. Fig. 12 summarizes the states of the control signals across the Gaussian pyramid.

State capacitors C_{pij_O1} in the first octave are the combination of MiM structures of M5-M6 metal layers C_{pij} with capacitors realized with transistors C'_{pij} in order to keep dynamic errors low, leading to C_{pij_O1} = 330 fF. Capacitors C'_{pij} are isolated from the SC network during LAMs read-out through signal ϕ_{read_net}, leaving C_{pij} = 200 fF for these functions (see Fig. 5). Exchange capacitors in the first octave are set to C_E = 38.5 fF and realized with transistors. According to Eq. (4), the state to exchange capacitors ratio yields $\sigma_{SC\_O1} = 0.48\sqrt{n}$ for the scales in the first octave, with n being the number of clock cycles. Such scales are built with blocks SC_A, SC_B and SC_C. Blocks SC_A run the two terms of the Gaussian kernel with NEWS connectivity through the switches that connect state capacitors within a given PE. The other two terms of the Gaussian kernel are executed with blocks SC_B or SC_C, correspondingly providing inter-PE connectivity of a given state capacitor with its neighbors. As an example, and as seen in Fig. 5, the state capacitor which results from merging C_{pij} with C'_{pij} into C_{pij_O1} within the the first octave is connected to its eastern and southern neighbors through SC_A within the PE, while their northern and western connections comprise blocks SC_B in PEs of α type, and blocks SC_C in PEs of β type. Finally, signals ϕ_1 and ϕ_2 in the basic cell of the double-Euler SC network of Fig. 2 are implemented with signals $\phi_{1\_O1\_pij}$ and ϕ_{2_O1} in blocks SC_A, ϕ_{1_pij} and ϕ_{2_O1O2} in blocks SC_B, and ϕ'_{1_pij} and ϕ'_{2_O1O2} in SC_C. $\phi_{1\_O1\_pij}$, ϕ_{1_pij} and ϕ'_{1_pij} are turn on to initialize C_E and C'_{pij} capacitors during image acquisition through CDS in every PE with signal ϕ_{read_net} high, as seen in Fig. 5 and Fig. 7.

The 1/4 downscaling from the first to the second octave occurs by shunting the four state capacitors C_{pij_O1} of the first octave with the 8 intra-PE exchange capacitors C_E, giving rise to larger state capacitors throughout the second octave as C_{pij_O2} = 4C_{pij_O1} + 8C_E for a given PE. In so doing, signals $\phi_{1\_O1\_pij}$ and ϕ_{2_O1} in blocks SC_A are always high in the second octave. Signals ϕ_{rw_pij}, ϕ_{rw_pij+1}, ϕ_{rw_pi+1j}, and $\phi_{rw\_pi+1j+1}$ are also high to shunt capacitors C_{pij} in the PE (see Fig. 5). Signals ϕ_1 and ϕ_2 in the basic cell of the double-Euler SC network of Fig. 2 are now given by the pairs ϕ_{1_pij} and ϕ_{2_O1O2}, and ϕ'_{1_pij} and ϕ'_{2_O1O2} in blocks SC_B and SC_C, respectively. Signals ϕ_{1_pij} and ϕ'_{1_pij} are used to initialize exchange capacitors for the second octave with blocks SC_B and SC_C. Also, as seen in Fig. 11, the NEWS connectivity for PEs of α type is given by two SC_B blocks along each direction. Similarly, two SC_C blocks along each cardinal direction are used for PEs of β type. This means that now the exchange capacitors for the second octave become $2C_E$. All in all leads to $\sigma_{SC\_O2} = 0.23\sqrt{n}$.

Finally, the 1/4 downscaling from the second to the third octave is carried out in two phases. During the first step the four state capacitors C_{pij_O2} of 4 PEs are shunted together through signals ϕ_{1_pij} and ϕ_{2_O1O2} high in blocks SC_B. Subsequently, these signals turn low, disconnecting PEs of β type from those of α type in every group of 4 PEs. As a consequence, the scales in the third octave are performed among PEs of β type through blocks SC_C, where ϕ'_{1_pij} and ϕ_{2_O3} play the role of control signals ϕ_1 and ϕ_2 in the basic cell of the double-Euler SC network of Fig. 2. Initialization of state capacitors is carried out with ϕ'_{2_O1O2} high. In this scheme, both exchange and state capacitors remain the same as in the second octave, so that σ_{SC_O3} = σ_{SC_O2}.

D. Peripheral Circuits

1) Gaussian Pyramid Read-Out: The Gaussian pyramid is read out through two frame buffers laid down at the top and bottom of the PE array, and labeled '1/2 frame buffer' in Fig. 3. Every register bank is assigned to the corresponding half of the PE array. The frame buffer split in two halves diminishes routing area.

Fig. 13(a) shows the 1/2 frame buffer. Every PE has two 8-bit registers assigned in the frame buffer, allowing the read-out and A/D conversion of two pixels at the same time. Such registers are named A and B in Fig. 13(b). Every frame buffer of the half PE array of 88 columns and 30 rows comprises 352 columns and 15 rows of registers. The 60 registers of a column of 30 PEs are placed in 4 columns of 15 rows each with the sequence $ABAB...$ of Fig. 13(b). As an example of read-out procedure, for the first column of PEs of the bottom half array- PEs across the 30th to the 59th row- the PEs from the 30th to the 44th row are A/D converted in column 0 in the register bank, while the PEs from the 45th to the 59th row are A/D converted in column 2 (both of them in reg. A in Fig. 13(b)). At the same time, the data converted in the previous cycle are read out of the chip in columns 1,3... (reg. B in Fig. 13(b)). Signal Reg_select allows selecting one of the two 8-bit registers, either A or B, yielding the A/D conversion. Finally, the 4-bit and a 9-bit row and column decoders are NOR MOS decoders with pull-up transistors.

The signal EoC from the in-PE comparator enables writing of the digital word generated by a global counter into the registers, which are implemented with an NMOS transistor at the input and a PMOS transistor in their feedback loop (Fig. 13(c)). The 8-bit register of a word includes a tristate at the output as showed in Fig. 13. The row decoder enables these tristates in a full row and all write the stored word in a per column vertical bus. Another tristate placed at the end of each column selects the column that must be read. The column tristate writes the data in the bus that drives the digital word to a buffer. This buffer reinforces and drives the 8-bit word to the output paths $digou$ and $digod$ (Digital Output Up/Down).

2) Analog Ramp and Voltage Bias Generation: The analog ramp for the 8-bit single-slope A/D converter is produced with an 8-bit current steering D/A converter [35]. The D/A converter is laid down at the left of the PE array in Fig. 3. The unity current for the D/A converter is set to 2 μA. The current from the D/A is converted to voltage in an external resistor. The D/A also comprises a 5-bit current steering to set up the offset of the ramp. Finally, the bias voltage generators of the gain amplifiers in the PE are implemented with wide swing transconductance amplifiers included on the left side of the die, within the block labeled 'Ana. Ramp' in Fig. 3.

IV. Experimental Results

A. Camera Module Prototype

Fig. 14 shows a camera module prototype composed of three interconnected boards. The first of them (carrier board) hosts the sensor chip (FPGP). The second board encloses an FPGA DEO-nano [36] to control the chip. The last one is a microPC (Raspberry Pi [37]) for visualization purposes. The optics is a C-mount type 35mm@f1/4 lens. The system is powered to 5 V through a plug $Jack/\mu USB$ type.

B. On-Chip Gaussian Pyramid

The chip operation depends on the value of the emulated Gaussian filter width, σ_{SC}. This is set during design through capacitors C and C_E with Eq. (4), where n stands for the number of clock cycles. Nevertheless, σ_{SC} may change during physical realization. Fig. 15 displays changes measured from the chip. The black line shows the designed σ_{SC} as a function of the number of clock cycles n. The blue line shows the σ_{SC} values of the scale-space extracted by iteratively comparing the outcome of the chip across the number of cycles n to an ideal scale-space $L(x, y, \sigma)$ on the image acquired by the chip through RMSE minimization. The red line is a polynomial fitted to the measured values. This experimental curve fits Eq. (4) by using exchange capacitor values of $C_E \approx 28$ fF and $C_E \approx 26.5$ fF for the first and second octaves, instead of the designed ones, i.e. C_E = 38.5 fF, due to tolerances and parasitics, which do not destroy chip functionality. It should be noted that both the exchange capacitors C_E, and part of the state capacitors C'_{pij} are implemented with transistors, while part of the state capacitors C_{pij} are MiM devices (see Fig. 5). Deviations among the experimental scales and scales designed with Eq. (4) are below 1% of the full scale, as it is illustrated by the right vertical axis in Fig. 15, where it is seen that the RMSE saturates around 2.5 in a scale of 255 (1% of FSO). Finally, Fig. 16 further illustrates the outcome of Gaussian filters realized by the chip by showing different scales obtained within the first octave.

C. Implementation Comparison

The chip generates a Gaussian pyramid of 3 octaves with 6 scales each in 8 ms. Time required for A/D conversion is included in this number. Thus, the chip can provide 125 digitally-encoded pyramids per second. Data conversion takes 200 μs per conversion and the clock cycle for the double Euler SC network is 150 ns. Relative energy consumption and throughput of our chip are 26.5 nJ/px at 2.64 Mpx/s.

Table III compares these metrics versus those provided by systems where Gaussian pyramids are obtained through digital signal processing following sensor read-out. Since some of these systems do not embed image sensors, energy for conventional CMOS imagers [38] scaled to the image resolution of the corresponding processor have been added for proper comparison.

Energy data in Table III do not include external memory accesses as they largely depend on the camera system. Their forecast would hence be inaccurate, and similar for all the Gaussian pyramid sensory-processing subsystems, including ours. Our chip is up to four orders of magnitude better than conventional and low-power MPUs in computer performance (Mpx/J), while the throughput is similar to that of the most efficient competitor.

Table IV further illustrates the performance of the chip versus other highly efficient sensory-processing CVIS chips with per-pixel circuitry. The chip in [6] performs 2D optic flow estimation. The PE array evaluates temporal contrast change by substracting two frames whose gains are set by a programmable gain amplifier. The chip in [42] runs 3 $\times$ 3 convolutions. The chip in [43] performs general purpose low-level image processing. Finally, the chip in [44] performs background subtraction. These functions are simpler than the generation of a Gaussian pyramid with 3-octaves@6-scales performed by the herein reported chip.

Still, the chips in [42] and [43] might compute Gaussian filters, as these are weighted convolutions. The metrics in Table IV correspond to isolated pairs of convolutions as Roberts or

Prewitt edge detectors, and to real-time edge detection at 25 fps, respectively. The evaluation of the Gaussian pyramid with these chips would certainly give different metric values, and it would require additional hardware to switch between octaves. The chip in [44] performs background subtraction with two digitally-programmable switched-capacitor low-pass filters per pixel. The energy overhead on our chip when compared to the chips in Table IV is partly explained by the higher complexity of the function that it runs. Differences in fill-factor and pixel pitch are also due to the larger complexity of our PE. Particularly, our chip and that in [6] embed an 8-bit single-slope A/D converter. Nevertheless, while [6] follows a per-column ADC architecture, our chip follows a per-pixel one to achieve full paralellism and hence large speed.

D. Application Assessment

The accuracy of the on-chip Gaussian pyramid has been assessed by incorporating hardware errors into the interactive tool reported in [45]. This tool employs the SIFT feature detector to perform visual tracking of six 2D textures on VGA-resolution videos. Visual tracking metrics are calculated along the application of homography, defined as the matrix that captures the transformation of the 2D textures from one frame to the next one; e.g. rotation.

Repeatability (RP) is the metric that we have calculated to assess the quality of visual tracking with the on-chip Gaussian pyramid [45]. As defined in [45], and formulated in Eq. (7), below, RP is the set of interest points S_{j-1} and S_{j-2} at frames $j-1$ and $j-2$ such that the geometrical distance between them after applying the corresponding homographies (H_{j-1} and H_{j-2}) from frames $j-1$ and $j-2$ to frame j are below a certain threshold normalized to the total number of interest points S_{j-1} or S_{j-2}. RP gives an estimate of the percentage of interest points whose allocation in successive frames is successfully forecast with the extracted homography.

$$RP = \frac{|(x_a \in S_{j-2}, x_b \in S_{j-1})| \, ||H_{j-2} \cdot x_a - H_{j-1} \cdot x_b|| < \epsilon}{|S_{j-1}|} \tag{7}$$

The RMSE values measured from the chip have been expressed as per-pixel local errors by finding the standard deviation of the normal distribution which corresponds to the given RMSE level. The normal distribution conveys the variability from chip manufacturing. These errors have been added to every scale of the Gaussian pyramid. Fig. 17 displays RP vs RMSE for RMSE of 0%, 1%, 2.5% and 5%. Our on-chip RMSE levels are below 1.2% of FSO. RP is the average of the aforementioned six 2D textures throughout all the frames of the corresponding videos with three different image transformations, namely, rotation, zoom and perspective distortion. The error bars, calculated as the standard deviation throughout the averaged data, reports RP degradations which are tolerable for most applications. In fact, as reported in [45], the temporal distance between consecutive frames has a larger impact on RP. In this regards, the large Gaussian pyramid calculation throughput of our chip becomes an important asset as it enables to reduce the baseline distance between consecutive frames.

V. Conclusion

This paper presents a proof-of-concept CVIS of 176×120 pixels for the parallel computation of the Gaussian pyramid with a double-Euler SC networks. Cutting PE area through smaller state capacitors of the SC network might the be the most straightforward way to upscale our architecture while keeping performance metrics. Eventually, a given resolution could not be met with a double-Euler SC network. In that case, resorting to a simple-Euler network might be a solution if the loss of accuracy is affordable for the targeted application framework. Measurements from our chip demonstrate that sensory-processing architectures with per-pixel mixed-signal processors outperform conventional architectures consisting of an imager and an MPU in terms of both energy consumption and throughput. Our results also show that unavoidable errors of the analog circuitry do not result into unfeasible Gaussian pyramids as it has been verified by visual tracking metrics with a publicly available image dataset. The main limitations posed by the type of SIMD-CVIS reported in this paper are direct consequences of the use of per-pixel circuitry and standard, planar technologies, namely: i) enlarged pixel pitch; and ii) reduced fill-factors. The former might constrain the use of this type of chips to applications where the object of interest is at a short distance to the camera. The latter calls mainly for applications with controlled illumination conditions. However, these limitations can be overcome by re-targetting our architecture into 3D vertically-integrated technologies, a task for which the circuits and methods reported in this paper can be re-used.

REFERENCES

[1] B. Roska and F. Werblin, ”Vertical Interactions Across Ten Parallel, Stacked Representations in the Mammalian Retina,” *Nature*, 410, pp. 583-587, 2001.

[2] T. Gollisch and M. Meister, ”Eye Smarter than Scientists Believed: Neural Computations in Circuits of the Retina”, *Neuron Review*, vol. 65, pp. 150-164, January 2010.

[3] C.L. Lee and C.C. Hsieh, ”A 0.8-V 4096-Pixel CMOS Sense-and-Stimulus Imager for Retinal Prosthesis,” *IEEE Transactions on Electron Devices*, vol. 60, no. 3, 1162-1168, 2013.

[4] J. Fernández-Berni et al., ”FLIP-Q: A QCIF Resolution Focal-Plane Array for Low-Power Image Processing,” *IEEE Journal of Solid-State Circuits*, vol. 46, no. 3, pp. 669-680, March 2011.

[5] S.J. Carey et al., ”A 100,000 fps Vision Sensor with Embedded 535GOPS/W 256 x 256 SIMD Processor Array,” *2013 Symposium on VLSI Circuits (VLSIC)*, pp. C182-C183, 2013.

[6] S. Park et al., ”243.3 pJ/Pixel Bio-Inspired Time-Stamp-Based 2D Optic Flow Sensor for Artificial Compound Eyes,” *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*, pp. 126-127, 2014.

[7] A. Rodríguez-Vázquez et al., ”ACE16k: The Third Generation of Mixed-Signal SIMD-CNN ACE Chips Toward VSoCs”, *IEEE Transactions on Circuits and Systems-I*, vol. 51, no. 5, pp. 851-863, May 2004.

[8] Anafocus Ltd. [Online]. Available. http://www.anafocus.com.

[9] R. Carmona-Galán et al., ”A Hierarchical Vision Processing Architecture Oriented to 3D Integration of Smart Camera Chips,” *Journal of Systems Architecture*, vol. 59, no. 10, Part A, pp. 908-919, 2013.

[10] T. Roska and A. Rodríguez-Vázquez, ”Towards the Analogic Visual Microprocessor”, ISBN: 0-471-95606-6, John Wiley & Sons, Chichester, 2001.

[11] V. Busskamp and B. Roska, ”Optogenetic Approaches to Restoring Visual Function in Retinitis Pigmentosa”, *Current Opinion in Neurobiology*, vol. 21, pp. 942-946, 2011.

[12] A. Torralba, ”How Many Pixels Make an Image?”, *Visual Neuroscience*, vol. 26, n. 01, pp. 123-131.

[13] M. Eldib, ”Behavior Analysis for Elderly Care Using a Network of Low-Resolution Visual Sensors”, *Journal of Electronic Imaging*, vol. 25, no. 4, pp. 041003-041003, 2016.

[14] Toshiba-Teli. http://www.toshiba-teli.co.jp/en/products/industrial/sps/sps.htm.

[15] A. Blug et al., ”Closed-Loop Control of Laser Power Using the Full Penetration Hole Image Feature in Aluminum Welding Processes”, *Physics Procedia*, vol. 12, pp. 720-729, 2011.

[16] E.R. Davies, ”Machine Vision: Theory, Algorithms, Practicalities”, Elsevier, 2004.

[17] P. Viola and M.J.A. Jones, ”Robust Real-Time Face Detection,” *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137-154, 2004.

[18] T. Tuytelaars and K. Mikolajczyk. ”Local invariant feature detectors: a survey,” *Foundations and Trends in Computer Graphics and Vision*, vol. 3, no. 3 pp. 177-280, 2008.

[19] J. Fernández-Berni et al., ”Bottom-up performance analysis of focal-plane mixed-signal hardware for Viola-Jones early vision tasks,” *International Journal of Circuit Theory and Applications*, vol. 43, no. 8, pp. 1063-1079, 2015.

[20] Á. Zarándy, ”Focal-Plane Sensor-Processor Chip,” Springer: Berlin, Germany, 2011.

[21] D. Lowe, ”Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60(2): 91-110, 2004.

[22] J. Fernández-Berni, R. Carmona-Galán, and A. Rodríguez-Vázquez, ”Low-Power Smart Imagers for Vision-Enabled Sensor Networks,” Springer Science & Business Media, 2012.

[23] M. Nathan et al. ”The Grasp Multiple Micro-UAV Testbed,” *IEEE Robotics & Automation Magazine*, vol. 17, no. 3, pp. 56-65, 2010.

[24] M. Suárez et al., ”A 26.5 nJ/px 2.64 Mpx/s CMOS Vision Sensor for Gaussian Pyramid Extraction,” *European Solid State Circuits Conference (ESSCIRC) ESSCIRC 2014-40th*, pp. 311-314, 2014.

[25] Jan J. Koenderink and A.J. van Doorn, ”Representation of Local Geometry in the Visual System,” *Biological Cybernetics*, vol. 50, no.5, pp. 363-370, 1984.

[26] T. Lindeberg, ”Scale-Space for Discrete Signals,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no . 3, pp. 234-254, Mar. 1990.

[27] P. Geissler, H. Haussecker, and B. Jähne. ”Handbook of Computer Vision and Applications,” *Vol. 2, Signal Processing and Pattern Recognition*. Academic Press, 1999.

[28] C.B. Umminger and C.G. Sodini. ”Switched Capacitor Networks for Focal Plane Image Processing Systems,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 2, no. 4, pp. 392-400, 1992.

[29] M. Suárez et al., ”Switched-Capacitor Networks for Scale-Space Generation,” *2011 20th European Conference on Circuit Theory and Design (ECCTD)*, pp.189-192, August 29-31, 2011.

[30] M. Suárez et al., ”CMOS-3D Smart Imager Architectures for Feature Detection,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 4, pp. 723-736, Dec. 2012.

[31] M. Suárez et al., ”Three Dimensional CMOS Image Processor for Feature Detection,” Patent No.: US 8,942,481 B2, Jan. 27, 2015.

[32] David A. Johns and Ken Martin, ”Analog Integrated Circuit Design”, John Wiley & Sons. Inc., 1997.

[33] A. El Gammal and H. Eltoukhy, ”CMOS Image Sensors”, *IEEE Circuits and Devices Magazine*, pp. 6-20, May/June 2005.

[34] Y.M. Chi et al.,”CMOS Camera with In-Pixel Temporal Change Detection and ADC,” *IEEE Journal of Solid-State Circuits*, vol. 42, no. 10, pp. 2187-2196, Oct. 2007.

[35] M. Suárez et al., ”In-Pixel ADC for a Vision Architecture on CMOS-3D Technology,” *2010 IEEE International 3D Systems*

Integration Conference (3DIC), pp. 1-7, 16-18 Nov. 2010.

[36] Terasic. [Online]. Available. http://www.terasic.com.tw/.

[37] Raspberry. [Online]. Available.http://www.raspberrypi.org/.

[38] Omnivision. [Online]. Available. http://www.ovt.com/.

[39] M. Murphy et al., "Image Feature Extraction for Mobile Processors". *2009 IEEE International Symposium on Workload Characterization*, pp. 138-147, 2009.

[40] Feng-Cheng Huang et al., "High-Performance SIFT Hardware Accelerator for Real-Time Image Feature Extraction," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 2, pp. 340-351, March 2012.

[41] G. Wang et al., "Workload Analysis and Efficient OpenCL-based Implementation of SIFT Algorithm on a Smartphone," *2013 IEEE Global Conference on Signal and Information Processing*, pp. 759-762, 2013.

[42] W. Jendernalik et al., "An Analog Sub-Miliwatt CMOS Image Sensor with Pixel-Level Convolution Processing," *IEEE Transactions on Circuits and Systems-I: Regular Papers*, vol. 60, no. 2, pp. 279-289, 2013.

[43] P. Dudek and P.J. Hicks, "A General-Purpose Processor-per-Pixel Analog SIMD Vision Chip, *IEEE Transactions on Circuits and Systems-I*, vol. 52, no. 1, pp. 13-20, January 2005.

[44] N. Cottini et al., "A 33 μW 64 $\times$ 64 Pixel Vision Sensor Embedding Robust Dynamic Background Subtraction for Event Detection and Scene Interpretation," *IEEE Journal of Solid-State Circuits*, vol. 48, no. 3, pp. 850-863, March 2013.

[45] S. Gauglitz et al., "Evaluation of Interest Point Detectors and Feature Descriptors for Visual Tracking," *Int. J. of Computer Vision*, vol. 94, pp. 335-360, 2011.

M. Suárez graduated in Physics in 2008 and received the Ph.D. in 2015 in the field of machine vision systems at the University of Santiago de Compostela, Spain. The research was done at Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS) in collaboration with the Institute of Microelectronics of Seville (IMSE-CNM-CSIC, Spain) through several stays along the PhD. He is the first inventor of an USPTO patent and received the 3rd Best Student Paper award at ECCTD 2013. Currently, he works as analog designer for Atomos GmbH in Germany.

V.M.Brea received his PhD in Physics in 2003 with honors. Dr. Brea has served in technical and steering committees of international conferences as ICDSC and CNNA. He has been awarded with the best and the third best student papers at ECCTD 2003 and ECCTD 2013, respectively. He has authored/coauthored around 100 papers in refereed journals, conferences and workshops. Currently he is an Associate Professor at Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS), University of Santiago de Compostela, Spain. His main research interests lie in the design of hardware architectures and CMOS solutions for computer vision, especially in early vision, as well as micro-energy harvesting.

J. Fernández-Berni received a B. Eng. degree in Electronics and Telecommunication in September 2004, a M.Sc. degree in Microelectronics in December 2008 and his Ph.D. in June 2011 with honors, from the University of Seville, Spain. From January 2005 through September 2006, he was working in the Telecommunication Industry. He has been a visiting researcher at the Computer and Automation Research Institute (Budapest, Hungary), Ghent University (Ghent, Belgium) and the University of Notre-Dame (IN, USA). Dr. Fernández-Berni has authored/co-authored some 50 papers in refereed journals, conferences and workshops. He is also the first author of a book and two book chapters as well as the first inventor of two licensed patents. He received the Best Paper Award in ”Image Sensors and Imaging Systems, SPIE Electronic Imaging 2014, San Francisco CA, USA” and the Third Prize of the Student Paper Award in ”IEEE CNNA 2010: 12th Int. Workshop on Cellular Nanoscale Networks and their Applications, Berkeley CA, USA”. His main areas of interest are smart image sensors, vision chips and embedded vision systems.

R. Carmona-Galán (M'04)(SM'16) graduated in Physics and got a Ph.D. in Microelectronics from the University of Seville, Spain. He worked as a Research Assistant at Prof. Chua's laboratory in the EECS Department of the University of California, Berkeley. He has been Assistant Professor of the Department of Electronics of the University of Seville. Since 2005, he is a Tenured Scientist at the Institute of Microelectronics of Seville (IMSE-CNM-CSIC). His main research focus has been on VLSI implementation of concurrent sensor/processor arrays for real time image processing and vision. He also held a Postdoc at the University of Notre Dame, Indiana (2006 - 2007), where he worked in interfaces for CMOS compatible nanostructures for multispectral light sensing. He has collaborated with start-up companies in Seville (Anafocus) and Berkeley (Eutecus). He has designed several vision chips implementing different focal plane operators for early vision processing. His current research interests lie in the design of low-power smart image sensors and 3-D integrated circuits for autonomous vision systems. He has coauthored more than 120 jorunal and conference papers and a book on low-power vision sensors for vision-enabled sensor networks. He is co-inventor of several patents. Ricardo Carmona-Galán is a Senior Member of the IEEE. He has been associate editor for IEEE TCAS-I and now is for Springer's Journal on Real-Time Image Processing. He got a Certificate of Teaching Excellence from the University of Seville. Very recently, he received the Best Paper Award of the IEEE-CASS Technical Committee on Sensory Systems at ISCAS 2015, together with Dr. Vornicu and Prof. Rodríguez-Vázquez.

D. Cabello (M'96) received the BSc and PhD degreesin Physics from the University of Granada, Granada, Spain, and the University of Santiago de Compostela, Santiago de Compostela, Spain, in 1978 and 1984, respectively. Currently, he is a Professor of Electronics at Centro Singular de Investigación en Tecnoloxías da Información (CiTIUS), University of Santiago de Compostela, Spain. He has been the Dean in the Faculty of Physics between 1997 and 2002, and the Head of the Department of Electronics and Computer Science between 2002 and 2006, both in the University of Santiago de Compostela. His main research interests lie in the design of efficient architectures and CMOS solutions for computer vision, especially in early vision.

Ángel Rodríguez-Vázquez (F'96) (IEEE Fellow, 1999) received undergraduate and PhD degrees in Physics-Electronics with several national and international awards, including an IEEE award. After different research stays in University of California-Berkeley and Texas A&M University he became a Full Professor of Electronics at the University of Sevilla in 1995. He co-founded the Institute of Microelectronics of Sevilla, under the umbrella of the Spanish Council Research (CSIC) and the University of Sevilla and started a research group on Analog and Mixed-Signal Circuits for Sensors and Communications. In 2001 he was the main promotor and co-founder of the start-up company AnaFocus Ltd and served as CEO, on leave from the University, until June 2009, when the company reached maturity as a worldwide provider of smart CMOS imagers and vision systems-on-chip. His research is on the design of analog and mixed-signal front-ends for sensing and communication, including smart imagers, vision chips and low-power sensory-processing microsystems. He has authored 11 books, 36 additional book chapters, and some 150 journal articles in peer-review specialized publications. He has presented invited plenary lectures at different international conferences and has received a number of awards for his research (the IEEE Guillemin-Cauer best paper award, two Wiley s IJCTA best paper awards, two IEEE ECCTD best paper award, one SPIE-IST Electronic Imaging best paper award, the IEEE ISCAS best demo-paper award and the IEEE ICECS best demo-paper award). He was elected Fellow of the IEEE for his contributions to the design of chaos-based communication chips and neuro-fuzzy chips. His research work got some 6,700 citations;he has an h-index of 43 and an i10-index of 133. He has always been looking for the balance between long-term research and innovative industrial developments. AnaFocus Ltd. was founded on the basis of his patents on vision chips and he participated in the foundation of the Hungarian start-up company AnaLogic Ltd. He has Eight Patents filed, three of which have been licensed to companies. He has served as Editor, Associate Editor and Guest Editor for different IEEE and non-IEEE journals, is in the committee of several international journals and conferences, and has chaired several international IEEE and SPIE conferences. He served as VP Region 8 of the IEEE Circuits and Systems Society (2009-2012) and as Chair of the IEEE CASS Fellow Evaluation Committee (2010, 2012, 2013, 2014 and 2015).

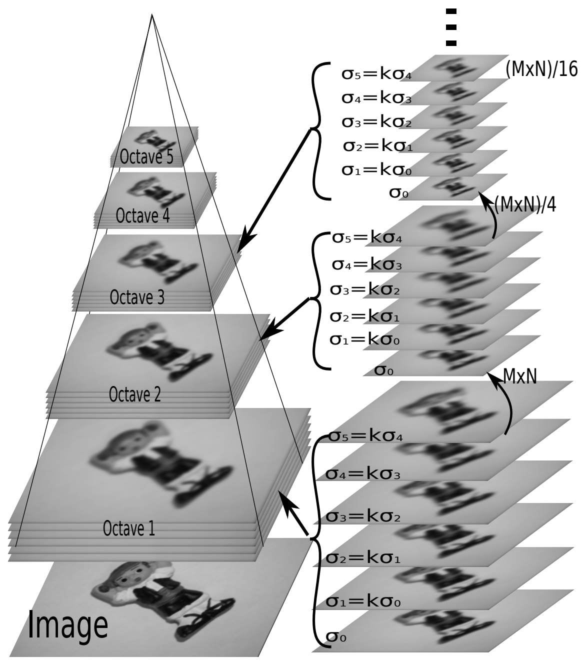

Fig. 1. Scale-space through the Gaussian pyramid with octaves and scales. Each octave has 1/4 the spatial resolution of the previous one, starting from the bottom. Thus, if the initial image has $M \times N$ pixels, images in the second octave have $(M \times N)$/4 and so forth.

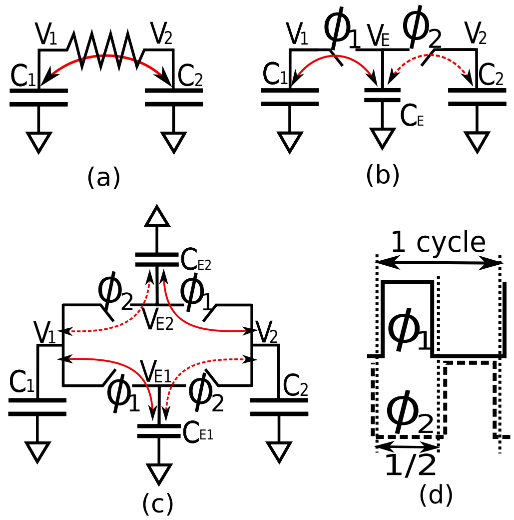

Fig. 2. Topologies for Gaussian filtering in 1D; (a) an RC network, (b) and (c) simple- and double-Euler SC networks, respectively. (d) non-overlapping control signals for SC networks.

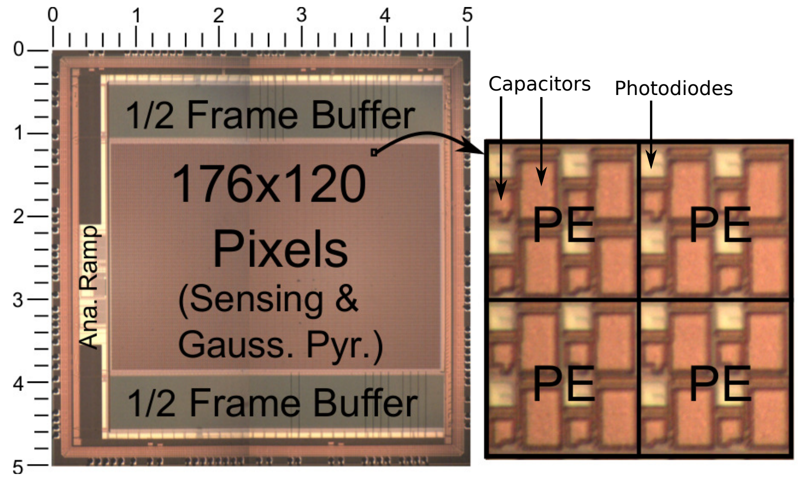

Fig. 3. Chip micrograph with dimensions (in mm) and a close-up of the PEs.

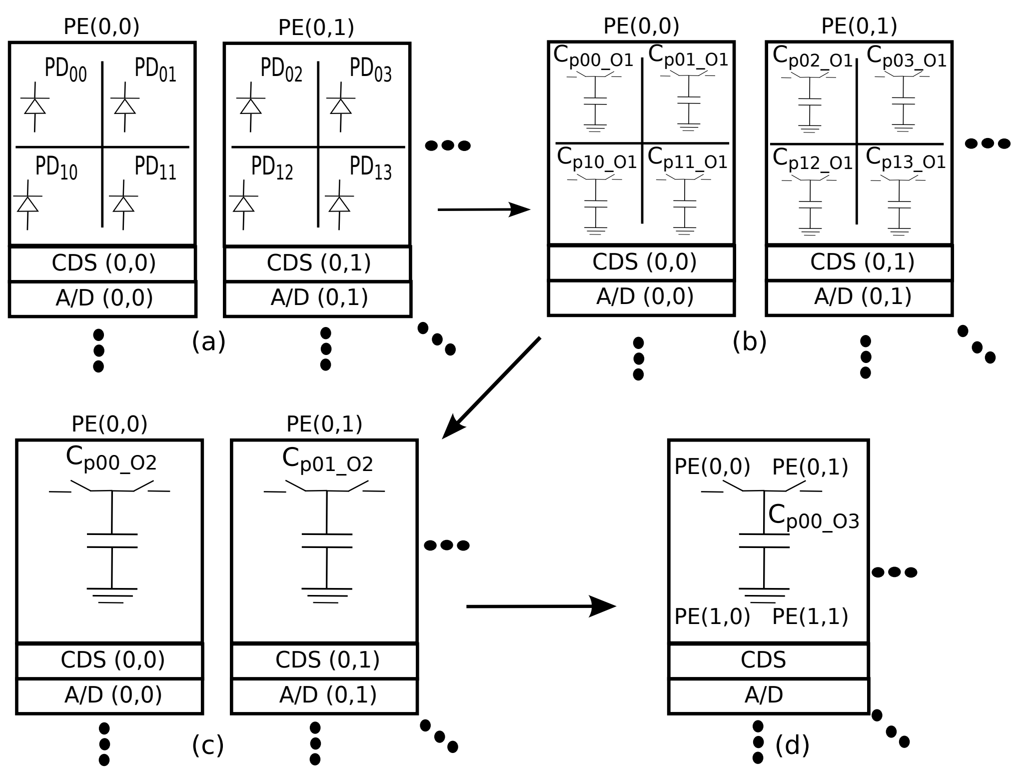

Fig. 4. PE array configuration across different functions of the chip: (a) image acquisition, where four photodiodes share one CDS and A/D converter in a PE, (b) first octave, where four state capacitors share one CDS and A/D converter in a PE (c) second octave, where four state capacitors in a PE are shorted together to perform downscaling, and there is a CDS and A/D per PE, and (d) third octave, where the state capacitors of 4 PEs are combined into only one to run downscaling.

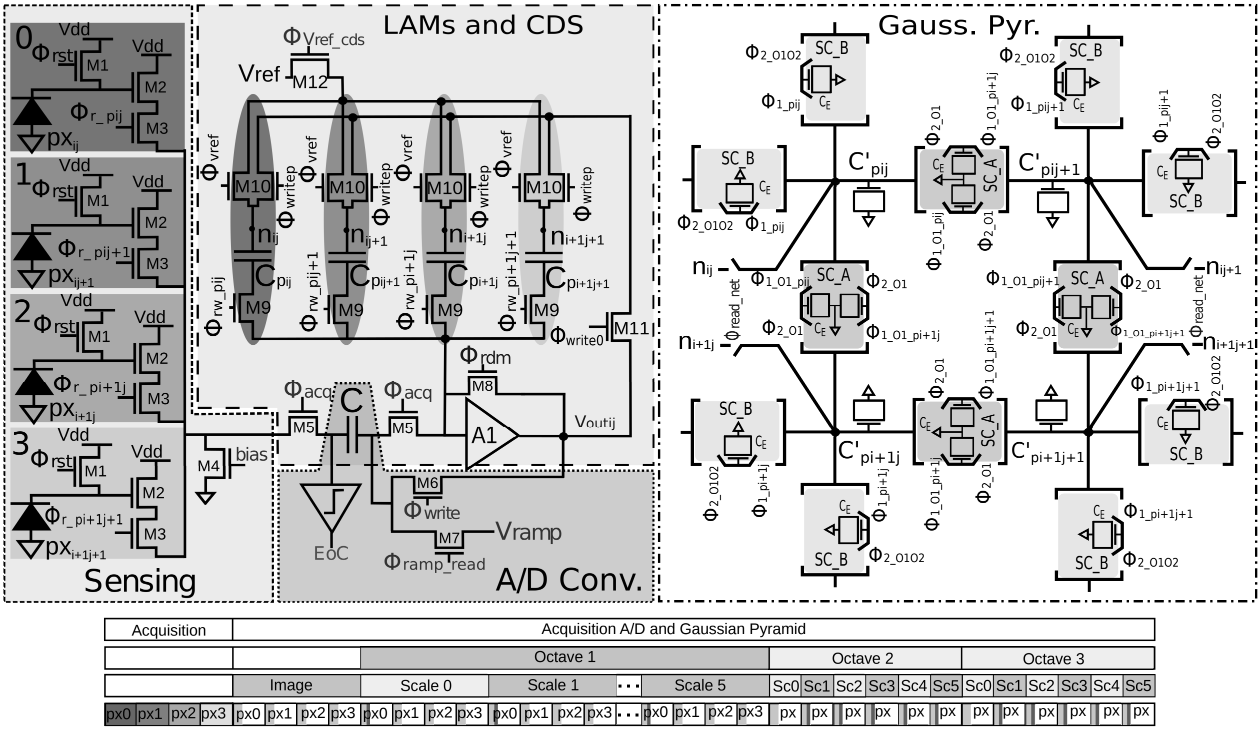

Fig. 5. PE and its associated time diagram. The PE is made up of four photosensors, four local analog memories (LAMs), one CDS circuit, one comparator for A/D conversion, and the local circuitry of the double-Euler SC network to build up the Gaussian pyramid.

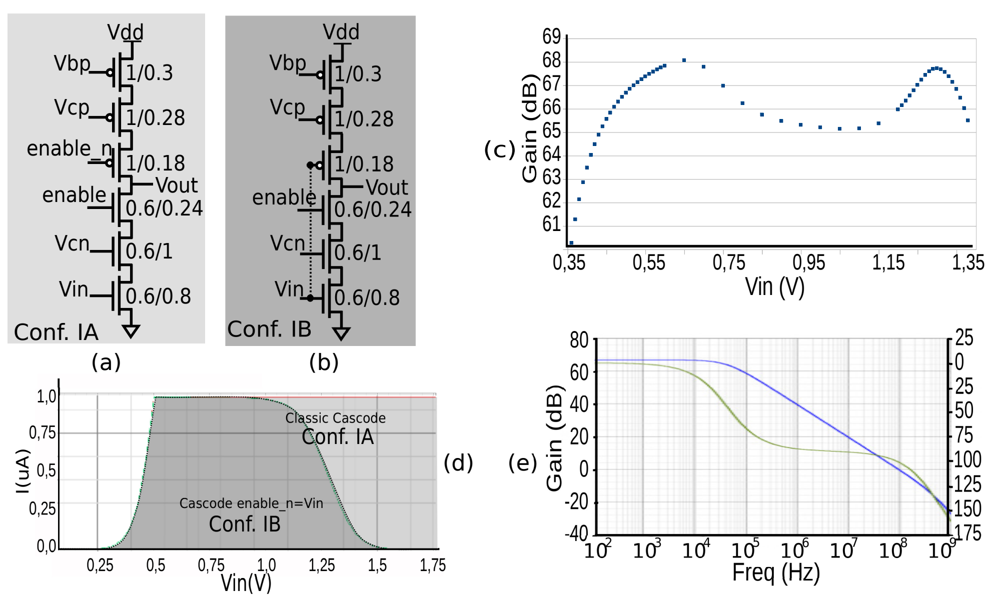

Fig. 6. Amplifier topologies used in the CDS, LAMs and comparator circuits of Fig.5 with some of their characteristics. (a) and (b) cascode configurations IA and IB. (c) gain versus input voltage. (d) current consumption vs input voltage in configurations IA and IB. (e) frequency response of configuration IA within the range of operation, [0.4, 1.3] V.

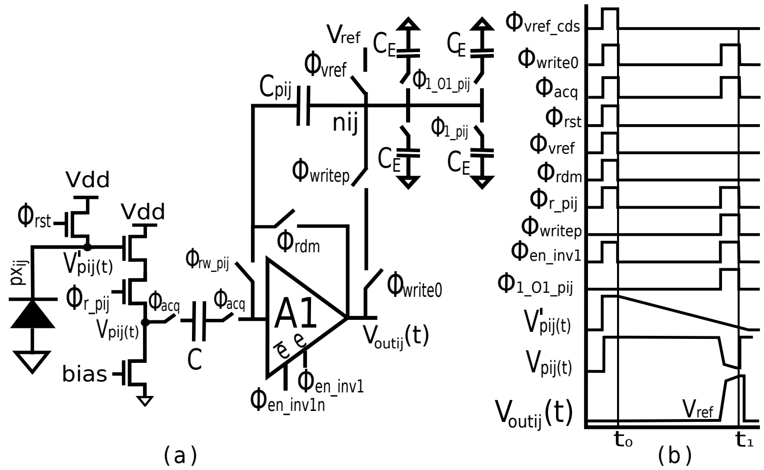

Fig. 7. Image acquisition through CDS on our chip. Signal ϕ_{rw_pij} selects the 3T-APS pixel associated with the position ij in the PE. Signal ϕ_{rw_pij} is high during the acquisition time for a pixel ij.

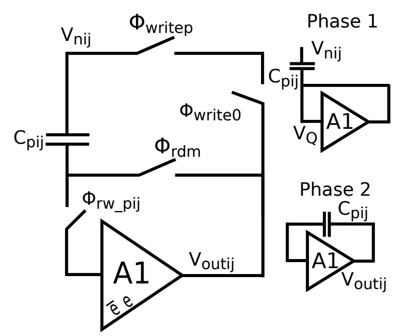

Fig. 8. LAMs working in two phases to store and read out scales across the Gaussian pyramid.

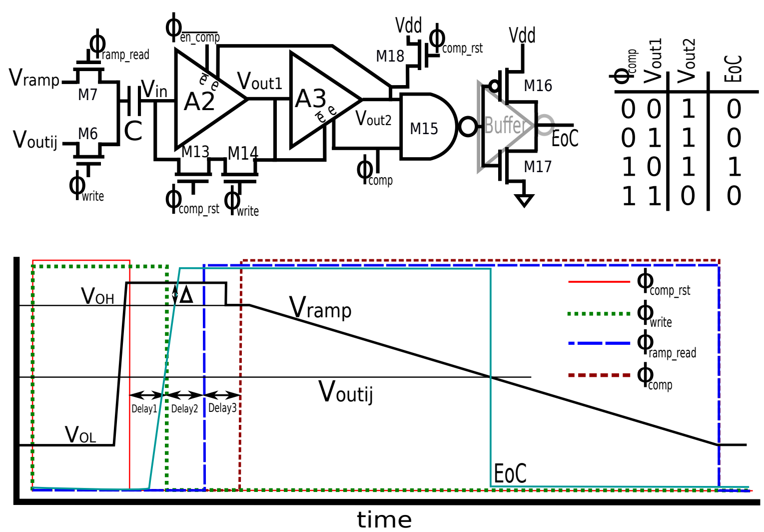

Fig. 9. Comparator of the in-PE 8-bit single-slope A/D converter with the time diagram of its control signals.

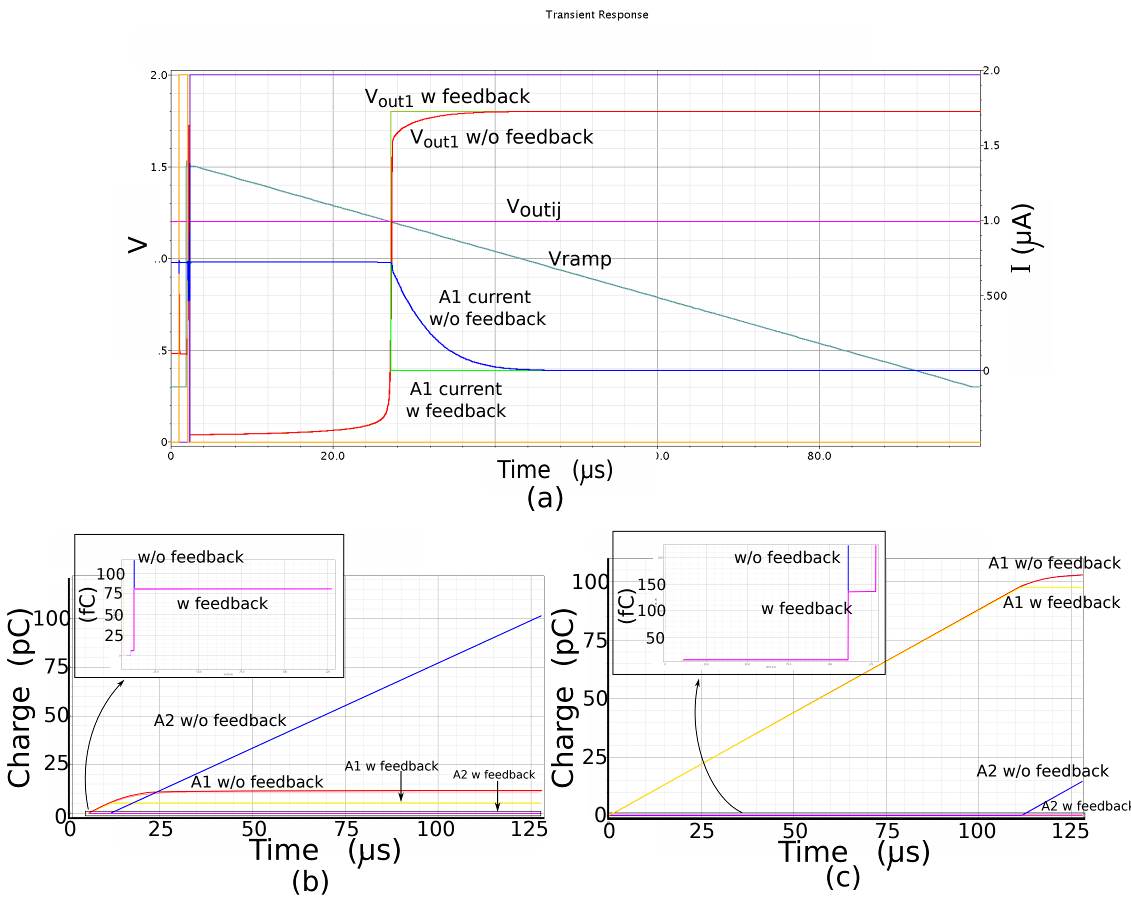

Fig. 10. Different waveforms of the in-PE comparator of Fig. 9 with (w) and without (w/o) feedback loop. (a) currents and voltages. (b) and (c) display currents integrated for input codes 250 and 40, respectively.

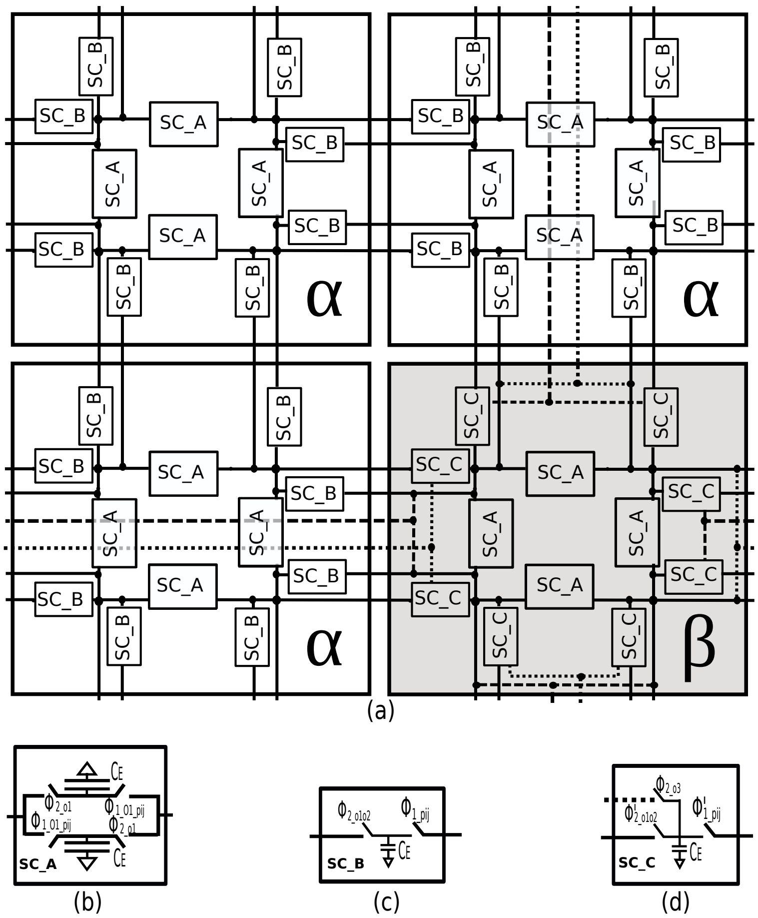

Fig. 11. (a) Double-Euler SC network for a grid of 4×4 pixels (2×2 PEs) of the chip. (b), (c) and (d) show the internal structure of blocks SC_A, SC_B and SC_C. Groups of 2×2 PEs comprise PEs of α and β type.

	Initialization	Scale	Downscaling
1st Oct.	$\Phi_{1\_O1\_pij} \rightarrow H$ $\Phi_{1\_pij} \rightarrow H$ $\Phi'_{1\_pij} \rightarrow H$	$\Phi_{1\_O1\_pij} \rightarrow S$ $\Phi_{2\_O1} \rightarrow S$ $\Phi_{1\_pij} \rightarrow S$ $\Phi_{2\_O1O2} \rightarrow S$ $\Phi'_{1\_pij} \rightarrow S$ $\Phi'_{2\_O1O2} \rightarrow S$	$\Phi_{1\_O1\_pij} \rightarrow H$ $\Phi_{2\_O1} \rightarrow H$
2nd Oct.	$\Phi_{1\_pij} \rightarrow H$ $\Phi'_{1\_pij} \rightarrow H$	$\Phi_{1\_pij} \rightarrow S$ $\Phi_{2\_O1O2} \rightarrow S$ $\Phi'_{1\_pij} \rightarrow S$ $\Phi'_{2\_O1O2} \rightarrow S$	$\Phi_{2\_O1O2} \rightarrow H,L$ $\Phi_{1\_pij} \rightarrow H,L$
3rd Oct.	$\Phi'_{2\_O1O2} \rightarrow H$	$\Phi'_{1\_pij} \rightarrow S$ $\Phi_{3\_O2} \rightarrow S$	

Fig. 12. State of control signals across the Gaussian pyramid. Symbols H and L refer to high and low states. H, L means that first the signal goes high, and subsequently low. All the former states are found during initialization or downscaling to change between octaves. Symbol S means that the signals are switching to generate the scales of the pyramid.

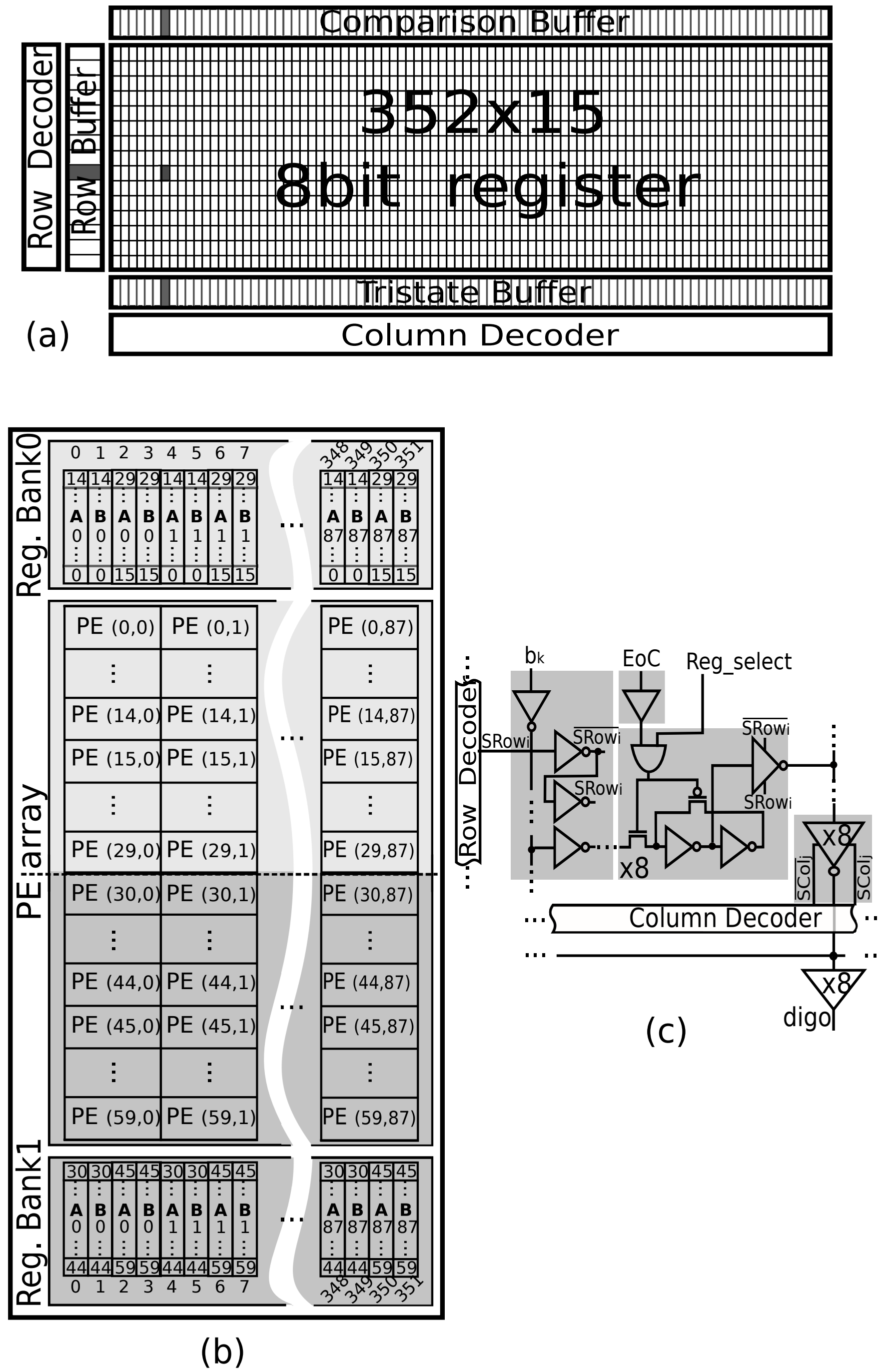

Fig. 13. 1/2 frame buffer bank for the single-slope A/D converter of half of the array is shown in (a). The PE-registers assignment is displayed on (b). The register circuitry can be seen in (c). EoC comes from the in-PE comparator. b_k is a bit of the digital word issued by an 8-bit global counter.

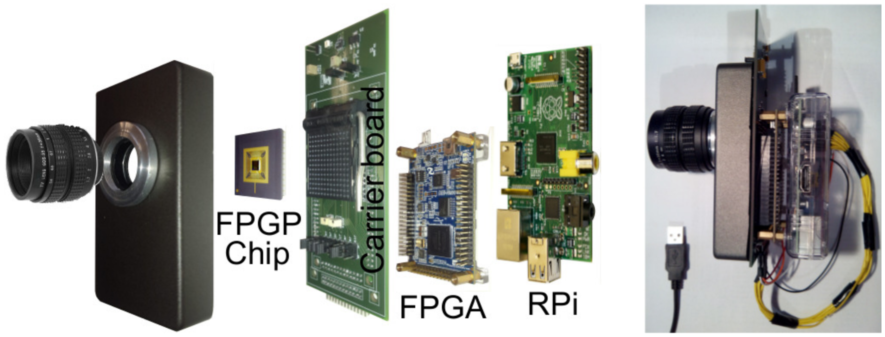

Fig. 14. Prototype camera module to extract the on-chip Gaussian pyramid.

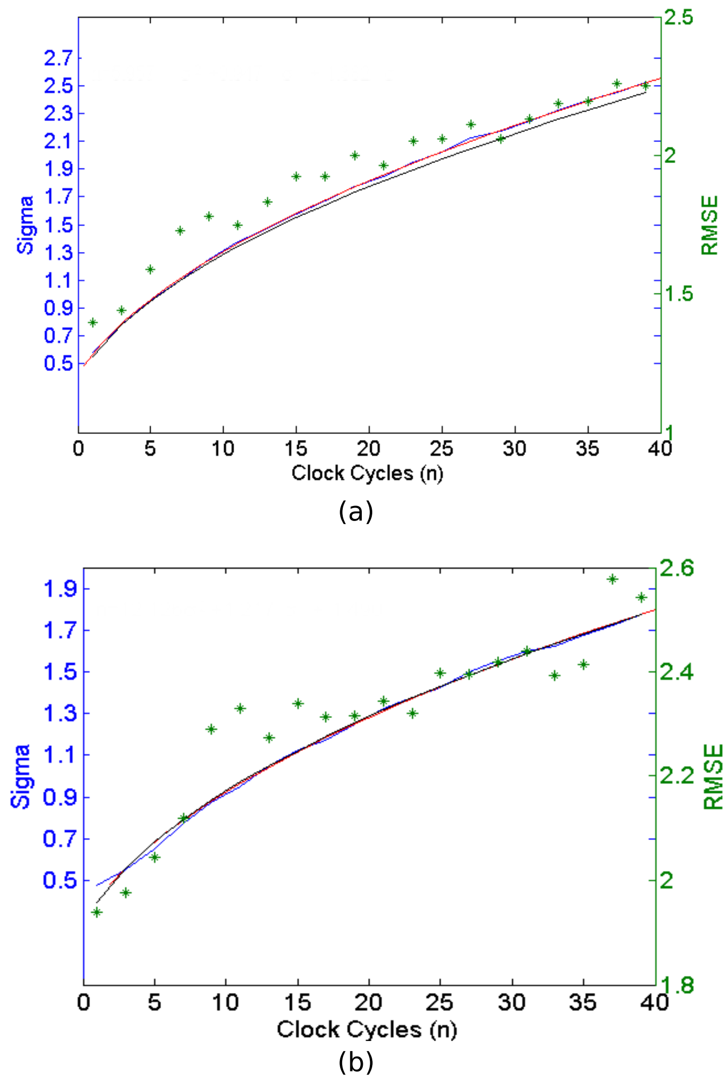

Fig. 15. On-chip σ_{SC} vs clock cycles n in the first and second octaves of the Gaussian pyramid.

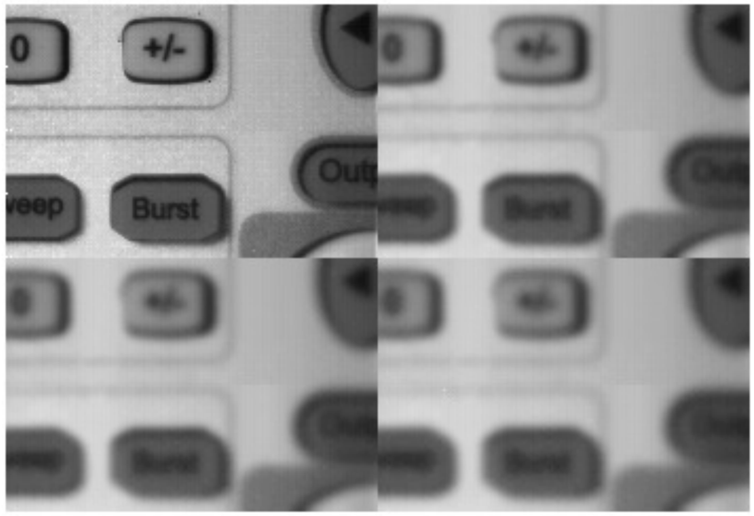

Fig. 16. Image acquisition and different snapshots of the on-chip Gaussian pyramid across the first octave. The upper left image is the input scene, the rest of the images from left to right and top to bottom correspond to σ=1,77 (clock cyles n=19), σ=2,17 (n=29), and σ=2,51 (n=39).

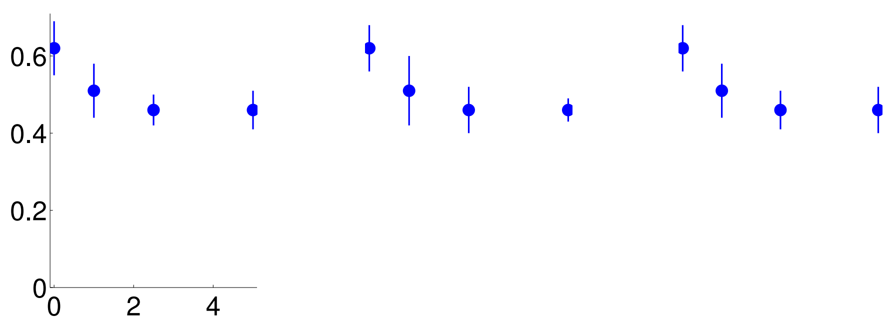

Fig. 17. Repeatability as a function of RMSE for three image transformations, namely, (a) rotation, (b) zoom and (c) perspective distortion.

TABLE I. PE TRANSISTOR SIZES (IN MICRONS).

	Width	Length		Width	Length
Photodiode	7.4	6.7	M1	0.24	1
M2	1.6	0.3	M3	0.24	0.6
M4	0.6	0.8	M5	0.24	1.4
M6	0.24	0.8	M7	0.24	1
M8	0.24	0.3	M9	0.24	0.2
M10	0.24	0.8	M11	0.24	0.2
M12	0.24	0.4			

TABLE II. COMPARATOR TRANSISTOR SIZES (IN MICRONS).

	W	L		W	L		W	L
M13	0.24	0.4	M14	0.24	0.8	M15	0.24	0.2
M16	2	0.2	M17	1.5	0.2	M18	0.24	0.2

TABLE III. COMPARISON OF OUR CHIP WITH CONVENTIONAL SOLUTIONS

HW Solution	Func.	Energy/frame	En./px (μJ/px)	Mpx/s	Mpx/J	Mpx/s.mm^2
This work 180 nm CMOS	Gauss. Pyr.	176 $\times$ 120 resol. 70 mW @ 8 ms 0.56 mJ/frame	0.027	2.64	37.7	0.11
Ref. [39] OV9655 + Core-i7	Gauss. Pyr.	VGA resol. 90 mW @ 30 fps + 35 W @ 136 ms 4.8 J/frame	15.5	2.26	0.064	0.007
Ref. [40] OV9655 + Core-2-Duo	Gauss. Pyr.	VGA resolution 90 mW + 35 W @ 2.1 s 73.7 J/frame	240	0.15	0.004	0.001
Ref. [41] OV6922 + Qualcomm Snapdragon S4	Gauss. Pyr.	350 $\times$ 256 resol. 30 mW + 4 W @ 98.5 ms 0.4 J/frame	4.4	0.91	0.23	–

TABLE IV. COMPARISON OF OUR CHIP WITH OTHER STATE-OF-THE-ART CVIS

HW Sol.	This work	Ref. [6]	Ref. [42]	Ref. [43]	Ref. [44]
Funct.	Gauss. Pyr. w A/D (SS 8 bits)	2D Optic Flow Est. w A/D (SS 8 bits)	3×3 Conv.	General Purpose Low-level Imag.-Proc.	Back. Subt.
Tech. & Res.	0.18 μm 176 × 120 px.	0.18 μm 64 × 64 px.	0.35 μm 64 × 64 px.	0.6 μm 21 × 21 px.	0.35 μm 64 × 64 px.
Fill-Fact.	10.25%	18.32%	23%	8.4%	12%
Pixel-Pitch	44 μm	28.8 μm	35 μm	98.6 μm	26 μm
En./px	26.5 nJ/px	0.89 nJ/px	0.19 nJ/px	0.52 nJ/px	0.62 nJ/px
Throughput	2.64 Mpx/s	0.49 Mpx/s	0.1 Mpx/s	0.11 Mpx/s	0.053 Mpx/s

Fig. 1. Scale-space through the Gaussian pyramid with octaves and scales. Each octave has 1/4 the spatial resolution of the previous one, starting from the bottom. Thus, if the initial image has $M \times N$ pixels, images in the second octave have $(M \times N)$/4 and so forth.

Fig. 2. Topologies for Gaussian filtering in 1D; (a) an RC network, (b) and (c) simple- and double-Euler SC networks, respectively. (d) non-overlapping control signals for SC networks.

Fig. 3. Chip micrograph with dimensions (in mm) and a close-up of the PEs.

Fig. 4. PE array configuration across different functions of the chip: (a) image acquisition, where four photodiodes share one CDS and A/D converter in a PE, (b) first octave, where four state capacitors share one CDS and A/D converter in a PE (c) second octave, where four state capacitors in a PE are shorted together to perform downscaling, and there is a CDS and A/D per PE, and (d) third octave, where the state capacitors of 4 PEs are combined into only one to run downscaling.

Fig. 5. PE and its associated time diagram. The PE is made up of four photosensors, four local analog memories (LAMs), one CDS circuit, one comparator for A/D conversion, and the local circuitry of the double-Euler SC network to build up the Gaussian pyramid.

Fig. 6. Amplifier topologies used in the CDS, LAMs and comparator circuits of Fig.5 with some of their characteristics. (a) and (b) cascode configurations IA and IB. (c) gain versus input voltage. (d) current consumption vs input voltage in configurations IA and IB. (e) frequency response of configuration IA within the range of operation, [0.4, 1.3] V.

Fig. 7. Image acquisition through CDS in our chip. Signal ϕ_{rw_pij} selects the 3T-APS pixel associated with the position ij in the PE. Signal ϕ_{rw_pij} is high during the acquisition time for a pixel ij.

Fig. 8. LAMs working in two phases to store and read out scales across the Gaussian pyramid.

Fig. 9. Comparator of the in-PE 8-bit single-slope A/D converter with the time diagram of its control signals.

Fig. 10. Different waveforms of the in-PE comparator of Fig. 9 with (w) and without (w/o) feedback loop. (a) currents and voltages. (b) and (c) display currents integrated for input codes 250 and 40, respectively.

Fig. 11. (a) Double-Euler SC network for a grid of 4×4 pixels (2×2 PEs) of the chip. (b), (c) and (d) show the internal structure of blocks SC_A, SC_B and SC_C. Groups of 2×2 PEs comprise PEs of α and β type.

Fig. 12. State of control signals across the Gaussian pyramid.

Fig. 13. 1/2 frame buffer bank for the single-slope A/D converter of half of the array is shown in (a). The PE-registers assignment is displayed on (b). The register circuitry can be seen in (c). EoC comes from the in-PE comparator. b_k is a bit of the digital word issued by an 8-bit global counter.

Fig. 14. Prototype camera module to extract the on-chip Gaussian pyramid.

Fig. 15. On-chip σ_{SC} vs clock cycles n in the first and second octaves of the Gaussian pyramid.

Fig. 16. Image acquisition and different snapshots of the on-chip Gaussian pyramid across the first octave. Image acquisition and different snapshots of the on-chip Gaussian pyramid across the first octave. (a) Input scene. (b) σ=1,77 (clock cyles n=19). (c) σ=2,17 (n=29). (d) σ=2,51 (n=39).

Fig. 17. Repeatability as a function of RMSE for three image transformations, namely, (a) rotation, (b) zoom and (c) perspective distortion.

Table I. PE Transistor Sizes (in Microns).

Table II. Comparator Transistor Sizes (in Microns).

Table III. Comparison of Our Chip with Conventional Solutions.

Table IV. Comparison of Our Chip with Other State-of-the-Art CVIS.